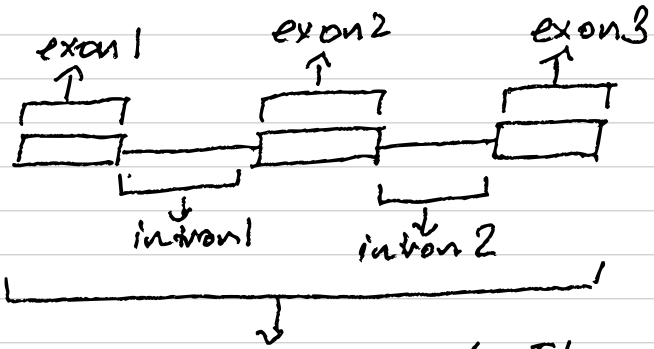
Task: Find the length of the introns between the exons in a transcript.



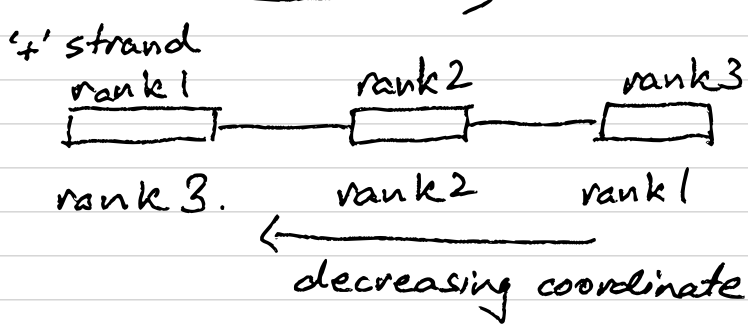
this is a transcript. It can have several exons interspaced with introns

In the data file provided

- transcript ID : the ID for the transcript.
- transcript name : the name of the transcript.
- sequence name : the chromosome ID
- start : the starting coordinate of an exon.
- end : the ending coordinate of an exon.
- width : the width of an exon = (end - start) + 1
- strand : + implies exons are ordered with increasing coordinate

  - implies exons are ordered with decreasing coordinates.

exon ID : ID for an exon.
exon name : name for an exon.
rank : position of the exon in the transcript.



Task 1: How many unique transcripts are there?

Task 2: How many unique exons are there?

Task 3: What is the average length of an exon? What is the median length?

Task 4: Find the length of the introns between the exons. (length must be a positive integer)

For the first exon in the transcript. intron length is 0

For the n^{th} intron in the transcript. intron length corresponds to the intron between the n^{th} and $(n-1)^{th}$ exon.

There is no intron after the last exon in the transcript. (to the right)

- Perform the above tasks once in R and once in python.
- Priority should be speed for Task 4

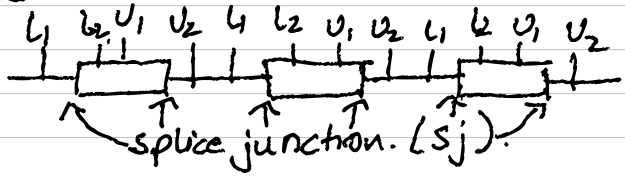
- Task 4 should have runtime of about 30 seconds or less.
- the code should be clean and well documented stating the procedure you used.

→ put the code and the dataset in a private github repository and add me as a collaborator:
username: talismanbrandi

→ commit message should be informative and relevant

Bonus task. (Both in python and R)

L_1, L_2, U_1, U_2 are coordinates



L_1 is 100 units of length before Sj left of an exon
L_2 is 100 units of length after Sj left of an exon
U_1 is 100 units of length before Sj right of an exon
U_2 is 100 units of length after Sj right of an exon.

The "L" regions and the "U" regions should not overlap. If the exon or intron is too short (< 200 units) 100 should be replaced by half the length of the exon or intron.

Note: l_1, l_2, u_1, u_2 should always be integers and not overlap. Floor or ceiling accordingly

Task: make 4 columns with l_1, l_2, u_1, u_2 for each exon. Remember, leftmost exon of a transcript will have $l_1 = 0$ and rightmost exon of a transcript will have $u_2 = 0$.

- if you have any questions or need hints please contact me.